



Standard Analytics Process- A step by step approach

Author: Fahad Bin Mahbub

Poster Advisor : Michael Lohle, Ph.D

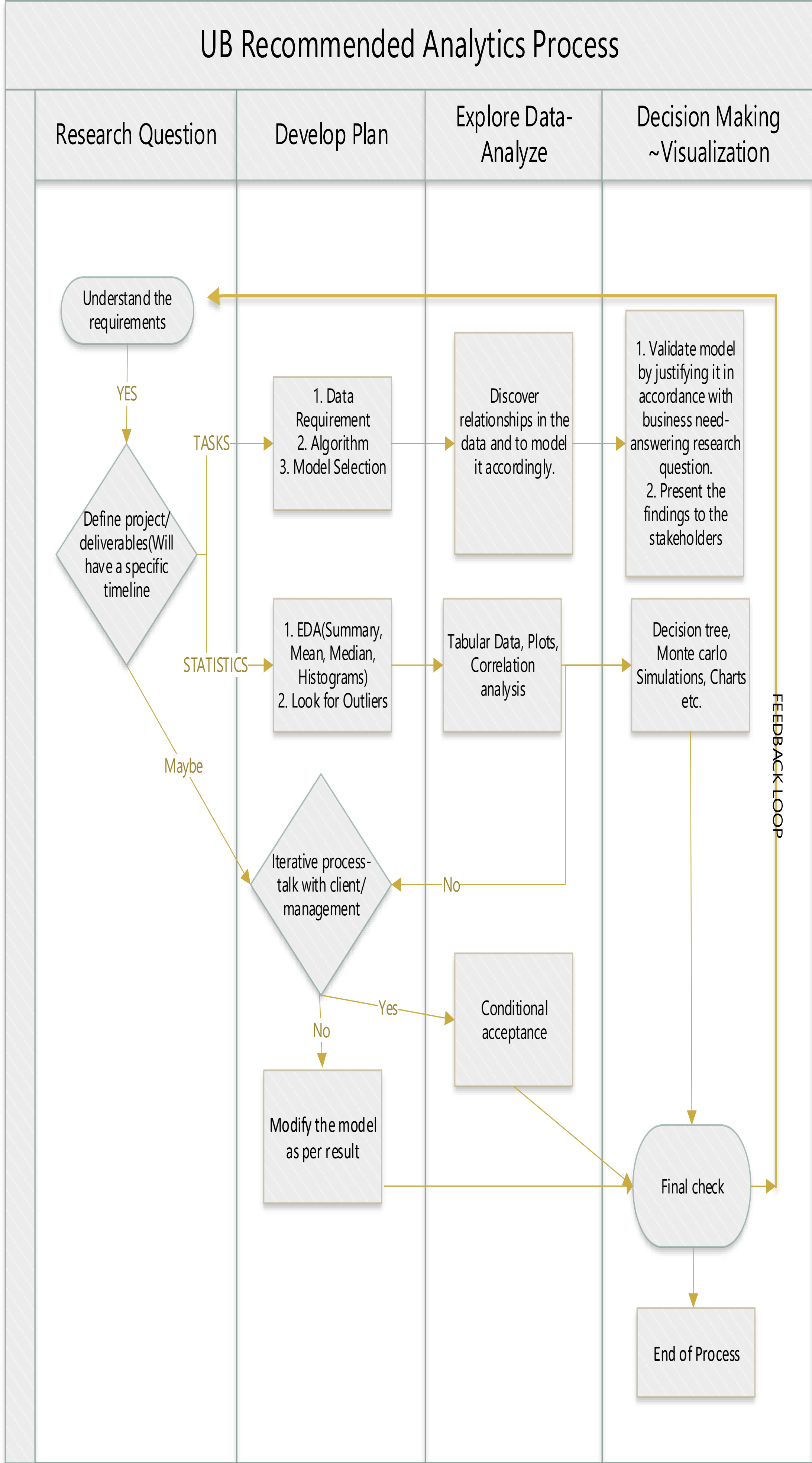
Department of Analytics and Systems, Trefz School of Business

University of Bridgeport

**Abstract**

UB’s analytics process is a step by step method that provides a template for how to approach and analyze data. It is also an example of how explicit knowledge can be documented and taught through process definition. Though the target audience for this process is business and data analysts, process details are provided to facilitate its application in other disciplines.

This analytics process has been mapped into four major segments: “research question,” “develop plan,” “explore data” and “decision making.” The four segments are broken into detailed steps. While variations to this process may be required in practice, this document depicts a standard process shared by the University of Bridgeport’s Analytics and Systems team at the Ernest C. Trefz School of Business. While an introduction to basic algebra and statistics provides a strong foundation for this process, business acumen is a must as everything starts with a business need. This “living document” is intended to be dynamic. Subsequent changes are welcomed pursuant to version control.



Steps	Description	Pre-Steps (Preparations)	To Do
1. Research question.....understand what the client/management wants	At this stage, one and only task is to 'ASK THE RIGHT QUESTION'. This is very important to start an analytical process.	Preliminary observation and understanding a client’s business model (how the business operates) is necessary.  Type A- “How can we use this data to help us with this decision?”  Type B- “What data can help us to make this decision?”  Type C- “How can this data help us/company?”	To formulate a research proposal or the question that needs to be asked one has to deep dive on the secondary data at this stage or to do some preliminary primary research (i.e.-focus groups)
2. Develop a program plan	Plan to include key decision steps and time frames.	Identify:  --Data requirements --Algorithms (Data types and structure) --Model selection methodology (based on Type of Data available) --Exploratory data analysis [EDA] ---How/What results will be presented.	Primary sources- Interview, Surveys. Secondary sources- Journals, Articles, Web etc.
3. Evaluation of quality and credibility of the chosen data/dataset	Assess the reliability and validity of the chosen data sources to address the business problem.		Determine: <ul style="list-style-type: none"><li>• Authority of the source</li><li>• When was the data published</li><li>• Purpose of publication</li><li>• Validation check as spelling, consistency, format, spacing etc.</li></ul>
4. Explore the data	This stage involves cleaning the data, making computations for missing data, removing outliers (to look up prior standards of the company for the same or similar experience), and transforming combinations of variables to form new variables.	1. Run summary statistics (mean, standard deviation, range, mode, median) to observe the distribution of the data for the credibility (Normal distribution/Or not?).  2. Create histograms, plot diagram (scatter plot, box plot) tabular data, basic visualizations like pie charts, bar charts, line charts, box plots.	After having an idea about the data; set ground to develop a hypothesis and related model.
5. Analyze the data	Discover the relationships in the data and to model it accordingly.	Run correlation analysis, regression (logistic/multiple) analysis. Often it is at this stage that the data is sliced and diced, and different comparisons are made while trying to derive actionable insights from the data	
6. Find the best solution (Optimize)	The analyst will select the optimal solution and model based on the lowest error.	Run several predictive models and select the best performing model based on its accuracy and outcomes.	R/Excel decision trees, random forests, Monte Carlo simulations are some great techniques to create and validate model(s).
7. Help management to make better decision			Convey findings (from the model) to management through communication and visual representations.
8. Visualization & presentation (Telling a story)	Step 1: Analyze the audience.  Step 2: Assess the outputs generated and confirm whether they will satisfy the business need; address the business problem in a format that will be palatable for the audience to understand.  Step 3: Draft the message.  Step 4: Test your message with peers prior to presentation.		Final presentation to the audience.
9. Document next steps, take action and establish and monitor feedback loop(s).	This avoids the ever present “analysis on an island” pitfall by ensuring clear actions are captured and analyzed, at which time the process begins again.		
10. Modify the model as required.		If necessary, run steps 1 and 2 for compatibility between the purpose of the analysis and findings derived from the analysis.	Revisit the research question to confirm that findings are representative of client’s and/or management’s need.

**Methodology**

UB recommended analytics process is an outcome of the cumulative effort of the faculty members at the department of analytics and systems for M.S. and MBA program. The author individually interviewed the faculties for a standard approach that should be considered as a learning experience for the students as and when they are given data sets to deal with.

The faculties provided their thoughts and year long practices for the author to accumulate and aggregate those to a meaningful process/steps which we have now as a standard process which is reasonable to follow.

**Conclusion**

This process is adaptive in nature and has controlled version. This flowchart and steps will be modified as we encounter new ways of analyzing data and the way we consider the word ‘analytics’. The author does not claim the work to be a conclusive one and the work encompasses the thoughts of all the faculties, not just a mere write-up.

The author believes that as UB develops its analytics program and we are exposed to more tools and complexities of handling data, this process/step by step approach will come handy to adopt, change and follow the way its intended. The author welcomes new inputs and modifications for the same.

Inserted: -

Class of Tools

Open source analytics tools

R  
Python  
Apache Spark  
Apache Storm  
PIG  
Google Analytics

Commercial analytics tools

SAS,  
SPSS,  
Excel,  
Matlab,  
Minitab, GA 360

Open source visualization tools

Candela  
Chartbox  
Polymaps  
Gephu

Commercial visualization tools

Tableau  
Qlikview  
Powerpivot  
Cognos

Open source database

Firebird  
Squirrel SQL  
EMS SQL Manager

Commercial database

Oracle  
MS Access  
PostgreSQL